



**caBIG**

*cancer Biomedical  
Informatics Grid*



## **Metadata Representation on The Grid: Requirements and Proposals**

George Komatsoulis, Denise Warzel, Ram Chilukuri

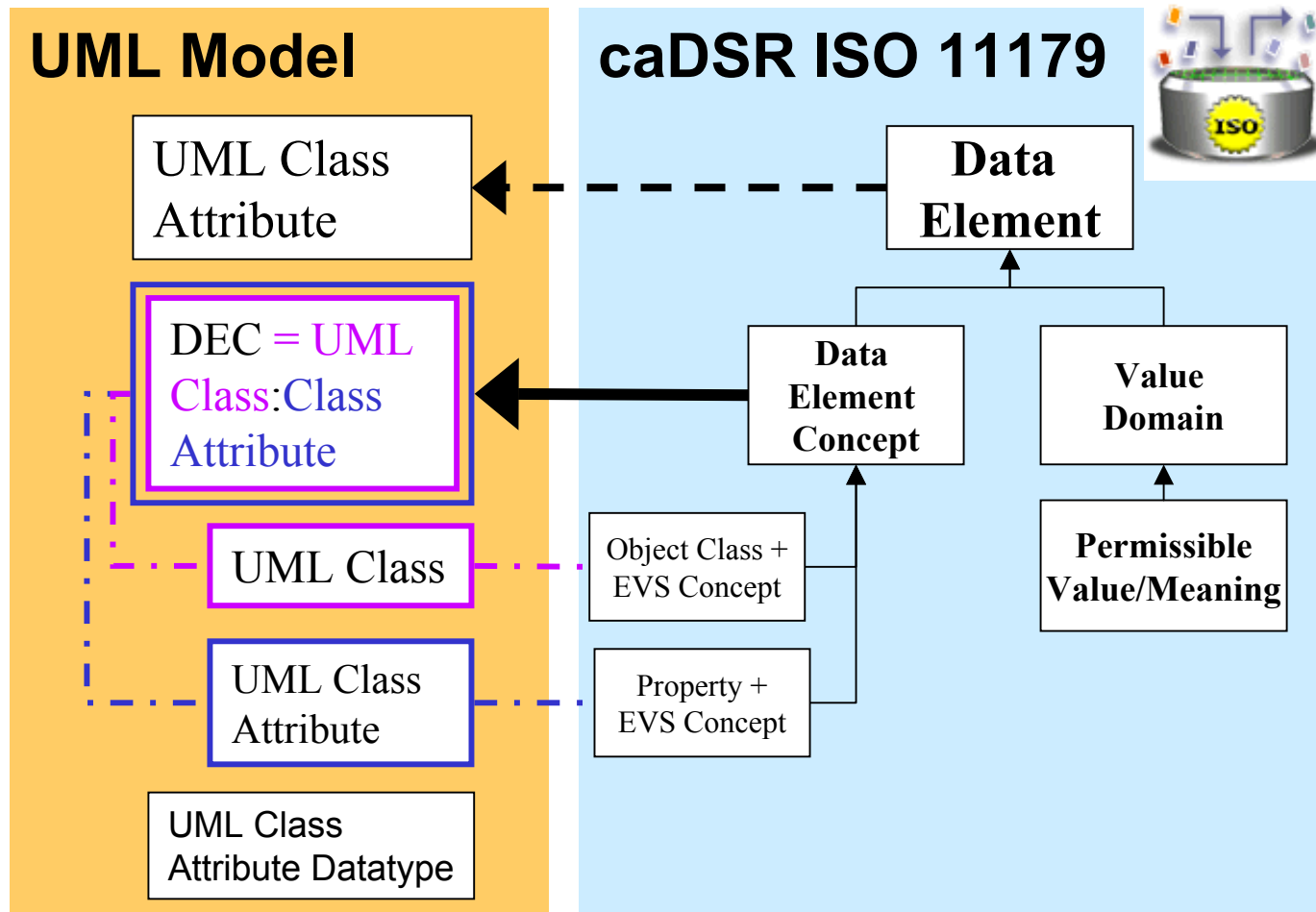
NCI Center for Bioinformatics

**10/25/2004**

# Agenda

- ▶ Brief Review of UML model to ISO11179/caDSR Mapping
- ▶ Metadata Use Cases
- ▶ Straw Man Model for Metadata Representation
- ▶ Open Issues

# Mapping UML Models to ISO11179/caDSR



# The Basic Metadata Use Case:

Retrieve enough metadata to:

1. Unambiguously interpret the input parameters required by a Grid resource.
2. Unambiguously interpret the meaning of a response from the Grid.
3. Successfully correlate/aggregate Grid results that represent either the same concept or subconcepts of a single superconcept.

Examples:

1. Given the results of a registry lookup, determine services that report data on Genes
2. Given a Discovery Query on a service, determine the types of analysis performed and the input parameters.
3. Given an XML describing an 'Agent' determine the meaning of the attributes included in the response.

## A 3-tier (Runtime) Metadata Model

- ▶ Tier 1: Data is formatted in XML corresponding to a UML model registered in the caDSR
- ▶ Tier 2: The structure of the data is constrained by an XML schema.
- ▶ Tier 3: The metadata is represented as an XML construct derived from the caDSR in some format (direct XML, RDF/RDF-S, OWL, etc.). We recommend RDF/RDF-S for its balance between flexibility and simplicity, and its ability to describe graphs.

# Examples

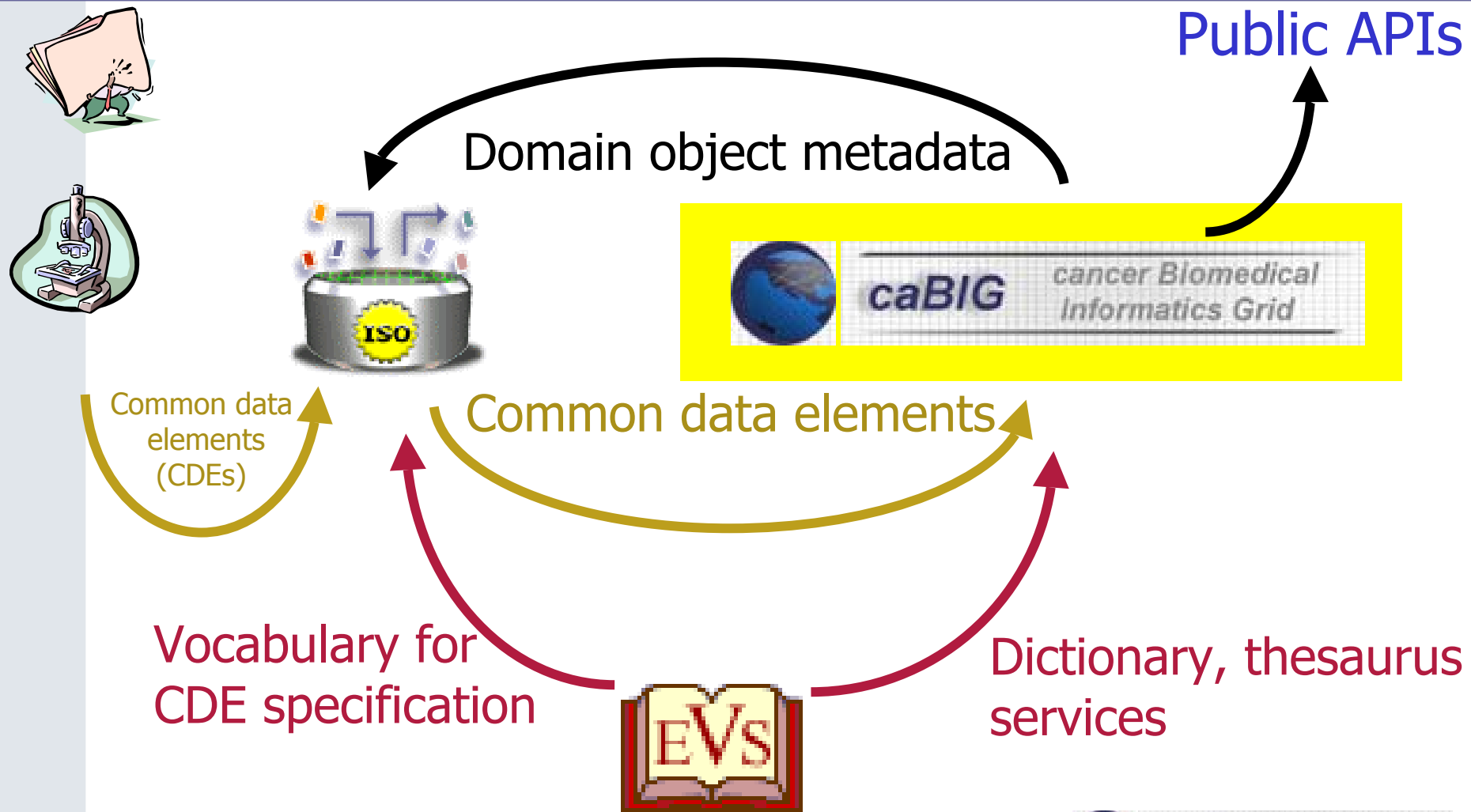
- ▶ A Data (i.e. Query) Service
- ▶ Discovery
- ▶ Advertisement

## Open Issues

- ▶ How do we link specific items in the XML schema to their metadata representations
- ▶ How do we expose large lists of enumerated values?
  - Constrain the values in the XML schema?
  - Retrieve the values from the Enumerated Value Domain of the caDSR extracted metadata
- ▶ How do we link the metadata to the EVS concepts? How should the concept information be presented?
- ▶ How do we compare/aggregate Data Elements when they are each based on up to 4 EVS concepts?

# Appendix: Supplementary Materials

# Infrastructure Wiring (Metadata Promise – caBIG Developer)



# caDSR Implementation of ISO/IEC 11179 Model

